Enhancing the stability of Organic Photovoltaics through Machine Learning

David, Tudur; Scapin Anizelli, Helder; Jacobsson, T. Jesper; Gray, Cameron; Teahan, William; Kettle, Jeffrey

Nano Energy

DOI:
https://doi.org/10.1016/j.nanoen.2020.105342

Published: 01/12/2020

Peer reviewed version

Enhancing the stability of Organic Photovoltaics through Machine Learning

Tudur Wyn David[a], Helder Anizelli[a], T. Jesper Jacobsson[b], Cameron Gray[a], William Teahan[a], Jeff Kettle[a]

a. School of Computer Science and Electronic Engineering, Bangor University, Dean St, Bangor, Gwynedd, Wales, UK, LL57 1UT
b. Young Investigator Group Hybrid Materials Formation and Scaling, Helmholtz-Zentrum Berlin für Materialen und Energie GmbH, Albert-Einstein Strasse 16, 12489 Berlin, Germany

Abstract

A machine learning approach for extracting information from organic photovoltaic (OPV) solar cell data is presented. A database consisting of 1850 entries of device characteristics, performance and stability data is utilised and a sequential minimal optimisation regression (SMOreg) model is employed as a means of determining the most influential factors governing the solar cell stability and power conversion efficiency (PCE). This is achieved through the analysis of the acquired SMOreg model in terms of the attribute weights. Significantly, the analysis presented allows for identification of materials which could lead to improvements in stability and PCE for each thin film in the device architecture, as well as highlighting the role of different stress factors in the degradation of OPVs. It is found that, for tests conducted under ISOS-L protocols the choice of light spectrum and the active layer material significantly govern the stability, whilst for tests conducted under ISOS-D protocols, the primary attributes are material and encapsulation dependent. The reported approach affords a rapid and efficient method of applying machine learning to enable material identification that possess the best stability and performance. Ultimately, researchers and industries will be able to obtain invaluable information for developing future OPV technologies so that that can be realised in a significantly shorter period by reducing the need for time-consuming experimentation and optimisation.

Keywords: organic photovoltaics, data-analytics, machine learning, stability, performance

1. Introduction

Organic Photovoltaic (OPV) development has been rapid over the past decades and the need to develop cheaper and more efficient renewable energy remains at the forefront of research priorities [1]. The nature of OPV research is such that a vast array of results has been

gathered, leading to data being generated with a plethora of materials and device structures. The body of work has led to substantial improvements in efficiency and stability, even if these still remain below those of silicon and other mature thin film technologies.

The body of available OPV data has now become large enough to enable the use of more advanced statistical analysis i.e. machine learning techniques, which potentially could provide insights beyond what is provided from individual studies. One of the key metrics for a solar cell technology's commercial potential is stability, which we in this study have explored by utilising machine learning (ML) techniques on a large dataset of OPV data. This allows interesting trends to be extracted from the underlying patterns within the dataset that goes beyond the standard approach of acquiring specific information by directly measuring the stability due to one or more changes. By deploying data analytical approaches in this manner, it is possible to determine which materials and stress factors have the greatest impact on the device stability. In addition, by acquiring an understanding of the materials and environmental attributes, which lead to more stable devices, the physical phenomena, which lead to degradation, can also be better understood. This represents a potential new paradigm in understanding OPV reliability.

Developing ML models for predicting efficiency and stability of OPV devices based on large datasets comes with a number of challenges. One of those are inconsistencies that arise from the fact that data comes from multiple authors working in different labs which introduces biases and random noise in the data. Others are inconsistencies in the set of materials that result in optimum stability, variability in material properties, and non uniformly employed testing protocols, with the result that the information concerning the relative impact of different layers upon the stability may be limited [2, 3]. A final challenge is that factors that have a significant impact on device performance simply is not adequately described or reported, or simply not yet properly known, and therefore it will not be possible to directly incorporate in the ML-models. One of the important questions that arises, and which the performance of the ML-models can answer, is to what extent we can understand and report the factors that really are important for device performance, and how much still is "dark knowledge" or just unknown.

In this work, we have utilised OV stability data which has been acquired from work undertaken between 2011-2019. This dataset contains data from **1850** devices, and is to date the most complete OPV stability dataset available. Using ML methods, we show that trends within this OPV dataset can be elucidated and models produced that enable OPV stability and performance to be estimated based on the device architecture and environmental testing

conditions. This methodology could be used to compare stability studies more fully and to analyse the relative significance of various environmental stresses and materials, and to ultimately identify key failure mechanisms in devices.

2. Methodology

2.1 Overview of data acquisition and ethics

The dataset used for this work was mostly obtained from the Danish Technical University (DTU) who ran the "lifetime predictor" on the Plastic photovoltaics website from 2011-2017 [4]. Additional papers between 2017 and 2019 were manually scraped. To source papers, 'Web of Science' was used to identify papers with OPV stability data. Device information and device data was subsequently extracted from the papers and added to the dataset. Information concerning device architecture and testing conditions was sourced. Device data was found in terms of performance and stability, which was extracted from tables or in figures; in which case a plot digitizer was employed in order to extract the relevant information from the figures.

Approximately 11% of the final dataset consists of data manually scrapped between 2017 and 2019. For all data added manually, the style and formatting was consistent with earlier dataset such that variability between the two subsets was minimised. In total, data from 1850 devices were used for this work. Papers were selected where device stability data is reported. This allows for parameters such as the initial efficiency (E_0), time taken for the device to reach 80% of the initial value (T_{80}) and time taken for the device to reach 80% of the stabilised value (T_{S80}) to be extracted. Whilst automated scraping using Python or R is possible, all the analysed data have been manually extracted from the papers.

2.2 Data acquisition and description of data format and categories

For the analysis methodology used in this paper, consistency in the format is essential and data has therefore been modified using the 'OSEMN' (Obtain, Scrub, Explore, Model, Interpret) process [5]. The "obtain and scrub" phase entails the acquisition and formatting of the data in a format suitable for the application of ML algorithms, such as a CSV file with consistency of format enforced throughout the dataset. In the case of the material/structural properties, a schematic of the metrics and generic structure of the OPV devices is illustrated

in FIG.1(a). Note that the active layer can comprise of both a single layer and a blend. This is represented by using two active layers in the dataset, represented by Active1 and Active2.

Initially, the appropriate attributes are acquired from each paper. The attributes for each device include the structure and materials, encapsulation and substrate type, test protocols, environmental conditions, light source. In excess of the device related attributes, there are also attributes related to measurement conditions such as temperature, light level, bias conditions and relative humidity. In total, the dataset we use contains 17 attributes, each of which contains a number of categories. Those are detailed in the supplementary information (SI-1). Given the number of attributes and the associated categories, the total dimensionality, i.e. the number of possible combinations of labelled device attributes, are 28725. That is far more than the 1850 combinations realised in the dataset, and even if many of those will be experimentally inaccessible, this illustrates the potential for extrapolation and suggestions for new promising combinations worth exploring experimentally that can came out of an analysis of this dataset.

In addition to the categorical features describing the devices, three lifetime and performance metrics have been extracted, namely E_0, T_{80} and T_{S80}, with T_{80} being focussed on primarily since, for OPVs, T80 is commonly used to assess the lifetime of these devices (see FIG 1(b)). In the authors' view, the T_{50} might be a more meaningful metric to use, but many papers do not report this or show graphs that cut off before this metric is reached, meaning that extrapolation is required, which can be subjective or misleading. In order to ensure consistency of reported data, all stability measurements correspond to tests performed under ISOS protocols. These standards are selected since most papers report according to these standards with ISOS-L and ISOS-D accounting for more than 96% of reports [4].

In any body of literature, the factors mentioned in each article may not be complete, with common factors being extensively explored and others being rarely investigated, resulting in a potentially uneven data distribution. However, this is mitigated to some extent for our specific analysis by the use of the large dataset of 1850 OPV devices. As researchers have investigated a diverse and complete set of factors in depth in a comprehensive manner, the natural consequence is that different factors have been explored which will naturally mitigated the imbalance problem for this research area. There are methods such as re-sampling [6] to improve the effectiveness of machine learning methods on unbalanced data. However, this was not warranted for our OPV dataset because, as stated, it was found to be relatively balanced except for some cases where there were found to be some materials which

occurred less than 5 times in the dataset. These entries were reclassified as being “Other” such that the distribution of instances was uniform for each attribute.

Once the datasets have been obtained, the ‘scrub’ phase can be implemented whereby, the data is cleansed and filtered, thus producing data which possess a consistent format. Therefore, careful inspection of the data is required in order to ensure that no unexpected values are present which could significantly affect the final result. For example, if the same materials are reported with slightly different names, or trade names, then these two quantities will be classified as two different features. At the scrub stage, the researcher needs to use their judgement to decide the level of detail they wish to apply ML for. For example, our approach could easily be used to allow groups to compare the stability of the same material from two different suppliers or identify the optimum thickness of a layer. We have not gone into this level of detail in this work and a full list of attributes is given in SI-1.

In addition to the data having an inconsistent format, some data may be missing or erroneous. There are several reasons as to why the dataset may contain missing values or contain errors made during the construction of the dataset or there could be a valid reason for having an empty cell in the dataset. For this work, we adopted a process called “imputation” which can be used to infer a missing value from the nature of the other attributes of the dataset [7]. There are various techniques to deal with missing data and values in machine learning such as using mean values [8], k nearest neighbours [9] and multivariate chained equation [10]. In this study, the data was cleansed and manual deductive imputation [11] was performed. The analysis also used the machine learning algorithm (SMOreg), whose implementation in WEKA of Alex J. Smola and Bernhard Scholkopf's sequential minimal optimization algorithm globally replaces all missing values. For example, considering the encapsulation or transport layer, a blank or missing entry could signify that this element was not present in the device. Therefore, inferring a categorical value of ‘none’ would be a suitable imputation for this quantity. In the case that the quantity cannot be inferred, due to lack of information concerning the device, the quantity “unknown” was imputed.

Subsequently, after the dataset has been formatted correctly, exploration and modelling can be implemented, whereby patterns and trends in the data can be investigated. The best ML algorithm can be chosen and applied for data analysis (discussed in section 2.3). An overview of the process adopted by us is shown in FIG. 2.

2.3 Machine learning approaches

Supervised learning has been adopted for this work because this allows a function to be fitted to the data in order to determine the response parameter (in this case T_{80}) based on the set of attributes. Subsequently, this allows the response parameter to be predicted based on combinations of attributes. Several algorithms which utilise supervised learning include support vector machine (SVM) [2], random forest [3] and decision trees [12]. Regression can be implemented using algorithms such as multilayer perceptron [13], principal component regressions [14] and sequential minimal optimisation regression (SMOreg) [15]. Throughout this work, SMOreg has been employed as the ML method. SMOreg has been selected since it produces an output model that can be analysed in terms of the weights of each attribute and, therefore, the significance of each attribute can be understood; this is ideal for this problem as we want to find which attributes have the biggest and least impact upon OPV stability. The weights of each attribute are obtained through the method of variational calculus using Lagrange multipliers to determine an optimum hyperplane, which separates the dataset into classes [16]. FIG.3 illustrates the classified data with a linear hyperplane separating the two classes, shown in blue and red. Greater detail concerning the theory and application of the SMOreg algorithm can be found in SI-2

The effectiveness of the SMO algorithm in characterising the data can be quantified by the correlation coefficient. The correlation coefficient (also known as Pearson Product-Moment correlation coefficient), r, determined in statistical regression analysis yields information regarding the linear dependence of one variable on the value of another [17]. The correlation coefficient can take values between -1 and 1 with 1 corresponding to a perfect positive correlation, whilst -1 corresponds to a perfect negative correlation. A value of 0 for the correlation coefficient signifies that there is no dependence of one variable on another. The strength of the correlation between two variables is given directly by the magnitude of r. Generally, a value of r between 0.5 and 1 is considered to be a strong positive correlation. The value of r can be calculated using equation (7).

$$r = \frac{\sum_i (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_i (x_i - \bar{x}_i)^2}\sqrt{\sum_i (y_i - \bar{y}_i)^2}} \quad (7)$$

In equation (1) x_i is the reference x value and $\bar{x}$ is the mean value of the reference values. Similarly, for y, the predicted response. In the equation above, $\sum_i (x_i - \bar{x})(y_i - \bar{y})$ corresponds to the product of the covariance of x and the covariance of y. The

term $\sqrt{\sum_i (x_i - \overline{x_i})^2}\sqrt{\sum_i (y_i - \overline{y_i})^2}$ corresponds to the product of the standard deviations of x and y respectively. We have used Waikato Environment for Knowledge Analysis (WEKA 3.8) for the ML modelling, utilising the built in .ARFF file viewer and ML algorithms.

3. Results and Discussion

3.1 Initial data exploration

FIG.4 shows the categorical violin plot for the full distribution of T_{80}, T_{S80} and E_0. These plots intuitively show the median, interquartile range and 1.5 × interquartile range. In addition, the plot shows the kernel density estimation to illustrate the distribution of each parameter. Analysis of this data allows the statistics of each of these variables to be determined as shown in TABLE.1. This demonstrates that the T_{80} and T_{S80} distributions are very similar with the near uniform distribution up to 10 days of stability, whilst E_0 displays a broad distribution for efficiencies of 3.5%. In addition, the interquartile range of the E_0 values is significantly less than for both T_{80} and T_{S80}.

3.2 Using the SMOreg algorithm for understanding stability data

Initially, the SMOreg algorithm has been applied to subsets of the data by considering three different testing conditions: the full dataset; data only conducted with light soaking (corresponding to ISOS-L test protocols), which relates to photostability; and data only conducted with thermal/damp-heat (ISOS-D test protocols), which relates to tolerance of oxygen, moisture, and heat. Our reasoning for doing this is that the failure modes in OPVs are heavily dependent on the testing conditions of the cells/modules under test. In order to effectively analyse the results from the ML algorithm, both the learning method, as well as the dataset, require optimisation. This has been performed in order to determine the training protocols that yield the highest correlation coefficient. This was achieved by training the ML algorithm using percentage split and cross validation.

Percentage Split is a re-sampling method that leaves out a random percentage of data in the training model for testing. This allows for us to assess the accuracy of the trained data based upon the test data. Cross-validation is a resampling procedure for training and testing an ML algorithm. Initially the dataset is reorganised and subsequently split into new groups, called “folds”. For each fold in turn, it is taken as a test set and the remaining folds as the training set. This is then repeated for each fold in turn. The model performance is then summarised for all iterations. In this way each entry in the dataset is assigned to one group, where it remains for the entire iteration.

The performance parameters are determined as a function of the proportion of the dataset used for training and as a function of the number of folds used for cross validation. For the dataset optimisation, this was achieved by varying the minimum value of each response parameter in each case. The results of this optimisation procedure are explained in greater detail in the supplementary information (SI-3). Low performing data has been removed such that noise within the dataset is minimised and, therefore, the algorithm does not overfit and make predictions based on this noise. In addition, by removing low performance devices, the algorithm will focus primarily on identifying the materials which lead to improved performance and stability.

The regression model predictions for the three subsets of data are shown in FIG.5. The algorithm performance metrics for each analysis are summarised in TABLE.2. The metrics indicate that the correlation coefficient improves when ‘subsets’ of data are analysed (i.e when data is separated into only ISOS-L data or ISOS-D data), rather than the data as a whole. The quality of the acquired model can be quantified in terms of the fitting parameters. The correlation coefficient can be employed to give a measure of this quality. In addition, the overfitting of the algorithm can be verified by comparing the correlation coefficients for training and testing; similar values for correlation coefficient for both training and testing are acquired, giving confidence that the model is not substantially overfitting. These are given in SI-3.

One of the benefits of using the SMOreg algorithm in this work, as stated, is that weightings for each attribute are obtained, which allows us to assess the relative impact each attribute has upon performance or stability. A positive SMOreg weighting corresponds to a positive influence on stability whilst a negative SMOreg weighting corresponds to a negative influence. Whilst the full list of weightings is given in SI-4, the 10 most beneficial and 10

most detrimental attributes for each subset of data is shown be found in TABLE.3. Yellow shading represents testing conditions. Green represents architectural components.

Several important features can be identified from the SMOreg weights. For the entire dataset, the attributes that most influence the stability positively are materials used in the first transport layer ('TL1') and the choice of active layer. In addition, the use of an LED light source is found to be beneficial for extending T_{80}. This is as expected due to the absence of UV light in the LED light sources. Significantly, all of the negative influences correspond to the ISOS testing protocol (and weighted by their relative severity) [18] and the light intensity. What one would expect *a priori* that harsher test conditions would have a negative impact on stability. This is also what we see, which provides confidence in our approach of using the SMOreg algorithm to evaluate how the attributes affect the stability. The effect of thermal cycling (ISOS-T-3) is found to be the most detrimental attribute, followed by the ISOS-L and then ISOS-D testing protocols. This highlights the importance of not only improving the material stability but also making the devices more robust against harsh environmental testing conditions.

Another factor used in the initial data analysis was the 'year of publication.' This showed a positive weighting for performance and stability; showing that more recent results in the dataset possessed better performance and stability. However, this factor was excluded from analysis henceforth as the primary aim of this paper is to report how the material and device impact stability.

When considering data only obtained using ISOS-L standards, the most influential attributes are also the structural components (i.e. substrate, transport layers, active layer, and electrodes), materials and configuration. The choice of light source was also found to play a significant role in improving the T_{80} lifetime.

When considering data only obtained from ISOS-D standards, the attributes that most positively and negatively affecting the degradation are found to be materials, architecture and the encapsulation method, illustrating the importance of protecting the device materials from environmental conditions during thermal or damp-heat testing. This demonstrates how the degradation during dark tests is dominated by the intrinsic stability of the materials used. The use of Tandem configuration leads to better dark stability (discussed earlier) along with the use of Perylenetetracarboxylic dianhydride (PTCDI) acceptor material and Bathophenanthroline (BPhen) as the top transport layer. Several interesting attributes can also be identified as being detrimental; the use of PEDOT:PSS, 'normal' device configuration and unsurprisingly, temperature.

To provide a summary of the data, TABLE.4 shows the three materials possessing the highest SMOreg weights for each layer within the device, when predicting T_{80} and considering 1) the full dataset, 2) the data from only ISOS-L tests and 3) the data from only ISOS-D tests. These top three weights illustrate the three most influential materials governing the stability and performance of the OPV devices, as predicted by the SMOreg algorithm.

Greater insight into the usefulness of ML approaches in elucidating trends and identifying key components can be gained from comparison of the highest weighted factors and the distribution of the T_{80} lifetimes for each attribute, shown in FIG.6. Inspection of the best electrode 1 components demonstrates that the three highest weighted materials are FTO, combined Cr and Al, and Ag. However, FIG.6 would suggest that these materials would not yield the greatest stability. Therefore, the ML algorithm has identified potential materials for stability enhancement which would not normally have been identified. This ability of the ML algorithm stems from the mechanism by which it finds the optimum hyperplane, whereby at least two Lagrange multipliers are simultaneously minimised. Inspection of the distributions for TL1, Active 1, Active 2 and TL2 illustrates the difficulty in manually identifying key components. Except for some high performing devices, very little variation in the T_{80} lifetimes can be identified. This is where the ML approaches are most useful; the most influential factors can still be extracted, as determined from the greatest SMOreg weights. The three materials identified for TL1 can be seen to correspond to experimental results which show significant stability. This is also true for the identified Active 1 and Active 2 materials, as well as the TL2 materials. This highlights the usefulness of this approach whereby significant features can rapidly and effectively be identified. If we look at the most stable devices found in that dataset, we do indeed see that the categories with high weights are used."

3.2.1 SMOreg oddities and biases encountered

Using a Tandem configuration plays the most significant part in the T_{80} improvement along with using ZnPc as an active layer material. The former is potentially a limitation within our dataset; there are only three reports in our dataset of stability testing using tandem configurations and, in all cases, the reported T_{80} value is high. Tandem devices are also primarily targeted by groups that already master the art of single junction devices. Despite the

thoroughness of our dataset, a small number of highly stable device will detrimentally bias the algorithm in favour of those architectures and testing conditions. This can occur when the data is asymmetrically distributed towards certain classes within attributes. For the highly stable tandem cells, the algorithm computes that the weightings of the normal or inverted configurations attributes to have a negative impact on the T_{80} time. In addition, testing in 'inert conditions' is found to be a detrimental influence for stability, which is highly contradictory of scientific evidence that oxygen and water could degrade materials. However, greater inspection of the dataset shows this is associated with poorer performing devices, which were tested in early stages of development and where the general foci of the paper is not on enhancing major gains in stability. The SMOreg algorithm will yield a large number of attribute weights, related to many of the features present in the dataset. Whilst the weightings to provide guidance on their impact on stability, discretion must still be exercised when identifying the most beneficial and detrimental features, such that the interpretation of the weights is meaningful.

3.3 Using the SMOreg algorithm for predicting initial stability, E_0

Whilst the focus of this work has been on stability data, the same methodology can be applied to understanding how different materials/architectures impact the efficiency prior to stability tests starting (defined as 'E_0'). The SMOreg algorithm has been applied to the dataset in order to predict E_0. The regression model prediction is shown in FIG.7 and the algorithm performance metrics can be found in TABLE.5. The 10 most influential and 10 most detrimental attributes have also been found and are listed in TABLE.6. The full list of weightings and the distribution of E_0 performances for each attribute can be found in SI-4.

Inspection of the weights for improving E_0 illustrates that all attributes are architectural components such as the active layer, the transport layers and the electrodes. Similarly, most of the attributes for deteriorating E_0 also correspond to architectural components. This provides a method of determining the optimum device architecture for maximised device performance.

It can be seen that the choice of active layer has the most significant impact on the value for E_0 along with using a Tandem configuration. The most commonly used polymer, P3HT, is determined to have a negative impact whilst other polymers such as PTB7 and PBDTTT-c display a positive impact. By performing the regression to predict E_0, the model

is being trained to understand which material combinations will result in higher efficiencies. However, care must be taken since the model does not, necessarily, know which combinations are impractical due to manufacture constraints. Therefore, interpretation of the acquired results still requires a certain level of discretion when drawing conclusions.

4 Challenges and perspectives

Historically, the great wealth of scientific knowledge was planned, recorded, and catalogued by generations of researchers in scientific articles [19]. The challenge faced today, in the era of digitalisation and rapidly progressing research, is the difficulty in processing these big data sets in an efficient manner. Applications of machine learning in the general area of material discovery is a rapidly growing area of research, with significant work being performed in the extraction of knowledge and information from past studies to allow predictions to be made years in advance of development [20]. We suggest that the methodology described here can be applied more widely in OPV and perovskite solar cell research, enabling the discovery of trends in material and device data, and provides guidance for new materials via fast screening of unexplored material and device combinations. Work in the application of this methodology is already becoming more prevalent. Association rule mining has been employed in order to determine factors leading to high efficiency as well as analysis of the heuristics for high efficiencies using decision tree classification [21]. In addition Thomas *et* al. employ ML techniques to optimise material composition and predict design strategies and performances of perovskite solar cells [22]. The stability of perovskite solar cells has also been analysed in [23] where the long – term stability of perovskite solar cells is investigated by determining the materials and methods that lead to stable perovskite solar cells. Therein, the effects of manufacturing methods, materials and storage conditions is analysed using association rule mining and decision trees. This allowed the optimum fabrication techniques to be highlighted as well as the optimum materials for each layer of the device, in order to minimise the degradation rates. In addition, studies on best practices for device testing and data reporting have been made conducted in order to acquire better quality data to be analysed using ML techniques [24, 25]. Such material/device-based inference methods can become an entirely new field of research at the intersection between AI/ML, natural language processing and material science, providing guidance regarding which aspects should be prioritised in order to achieve the biggest overall enhancement in device stability.

Our experience in this work highlights some challenges that lie ahead. The success of our supervised approach can partly be attributed to the choice of the data acquired. However, vastly different performance metrics can be obtained when considering the same structures and materials. Key to overcoming this challenge is ensuring publications report to the right level of detail and consistent reporting in terms of material details, test conditions and material/device processing. The recently published ISOS consensus standards provide some discussion about this and state that for approaches set out in the paper to be a greater future success, standardised reporting is key [26]. The importance of training data selection was demonstrated by comparing the 'entire dataset' with the data only obtained from 'ISOS-L testing'. We showed that selection of subsets of data improves the models trained on the dataset.

One of the questions addressed in the introduction is to what extent we understand and report the factors that really are important for device performance, and how much still is "dark knowledge" or just unknown. The attributes and categories we used in the modelling represent a large portion of what commonly goes into describing OPV devices when described in the literature. There are more factors we could point out that reasonably are of importance, e.g. layer thickness, synthesis temperatures, chemical providers, etc. The problem with many of those parameters are twofold. One is that they are often not reported. Another is that with too many categories, the number of devices for which the data is available is too small to properly train such a model.

The models we present here have an accuracy of 40.6% and 44.4% with respect to T_{80} and E_0 prediction, respectively. This is far from perfect and other phenomena which improves device performance could thus be attributed to other factors. Given the difference in performance of similar devices made at different labs a substantial, but yet unquantified, part of the discrepancy between the model and the reality could be attributed to either "dark" unreported knowledge or to hidden variables that are unknown. If we look into the future where we can expand the training data set with data for more devices and with a more fine grained meta-data mechanism, we will likely be able to elucidate more of those parameters.

Conclusion

Machine learning has been applied to a comprehensive OPV dataset as a rapid and effective screening technique for identifying the primary attributes in OPV degradation which have a positive or negative impact on the initial efficiency as well as the T_{80} lifetime. Supervised

learning and regression have been implemented using the SMOreg ML algorithm and has allowed the optimum material for each device layer to be identified as well as highlighting the role of the environmental conditions on stability and performance by considering the T_{80} and E_0 response based purely on the device architecture and testing conditions. The SMOreg algorithm has been applied to the dataset separated for tests conducted under both ISOS-L and ISOS-D testing compatibilities for E_0 and T_{80}. By separating the data in such a way, the role of the different materials and stress attributes on both the operational stability as well as the intrinsic material stability can be separately identified. For ISOS-L testing, the choice of light spectrum as well as several active materials were found to be vital in order to improve the stability. In contrast, for tests conducted under ISOS-D, the primary attributes affecting the stability were predominantly material dependent along with the use of encapsulation.

From inspection of the most promising materials and device architectures identified via this methodology, it is hoped that future studies will build on these findings in order to experimentally verify and test these predictions. In addition, the materials identified herein will provide inspiration for other research groups to develop and enhance the stability of their devices.

Scientific progress relies on the efficient assimilation of results from a plethora of research outputs in order to choose the most promising way forward and to minimise re-invention. The capability of identifying the best material combinations and the role of the testing condition in the performance and stability of OPV devices is crucial for identifying the priorities of research. In addition, classes of materials that are more robust against environmental stresses can be identified whilst classes of materials with lower stabilities can be avoided. We hope that this work will pave the way towards making the vast amount of information found in scientific literature accessible to individuals in ways that enable a new paradigm of machine-assisted scientific breakthroughs.

Acknowledgements

The authors would like to thank Suren Gevorgyan for sharing the dataset prepared on work between 2011-2017 as well as his support for organic photovoltaic group. We would like to thank Prof Fred Krebs, Dr Gevorgyan and others for establishing this datset and for whom this work wouldn't have been possible. We would also like to thank Rico Meitzner of the Friedrich-Schiller University of Jena, Simon Zuefle of FLUXiM AG and Priyanka Tyagi for their insightful discussions. JK and TWD have been supported by the Solar Photovoltaic

Academic Research Consortium II (SPARC II) project, gratefully funded by WEFO. This ppublication is also based upon work from COST Action CA16235 PEARL PV supported by COST (European Cooperation in Science and Technology) see www.cost.eu and www.pearlpv-cost.eu/ for more details.

References

[1] S. Michas, V. Stavrakas, N. A. Spyridaki, A. Flamos, **2019**. Identifying Research Priorities for the further development and deployment of Solar Photovoltaics. *International Journal of Sustainable Energy*, **2019,** *38*, 3, pp.276-296.

[2] C. Cortes, and V. Vapnik, **1995**. Support-vector networks. *Machine learning*, *20*, 3, pp.273-297.

[3] L. Breiman, **2001**. Random forests. *Machine learning*, **2001,** *45*, 1, pp.5-32..

[4] T. W. David, H. Anizelli, P. Tyagi, C. Gray, W. Teahan, J. Kettle, **2019**. Using Large Datasets of Organic Photovoltaic Performance Data to Elucidate Trends in Reliability Between 2009 and 2019. *IEEE Journal of Photovoltaics*, **2019,** *9*, 6, pp.1768-1773..

[5] H. Manson, C. Wiggins, "A Taxonomy of Data Science", http://www.dataists.com/2010/09/a-taxonomy-ofdata-science/, accessed: April 2020.

[6] H. He, E. A. Gracia, **2009.** Learning from imbalanced data. *IEEE Transactions on knowledge and data engineering*, **2009,** *21*, 9, pp. 1263-1284.

[7] J. L. Schafer, **1999**. Multiple imputation: a primer. *Statistical methods in medical research*, **1999,** *8*, 1, pp.3-15.

[8] Little, R.J. and Rubin, D.B., 2019. *Statistical analysis with missing data* (Vol. 793). John Wiley & Sons.

[9] Zhang, S., 2012. Nearest neighbor selection for iteratively kNN imputation. *Journal of Systems and Software*, *85*(11), pp.2541-2552.

[10] Buuren, S.V. and Groothuis-Oudshoorn, K., 2010. mice: Multivariate imputation by chained equations in R. *Journal of statistical software*, pp.1-68.

[11] De Waal, T., Pannekoek, J. and Scholtus, S., 2011. *Handbook of statistical data editing and imputation* (Vol. 563). John Wiley & Sons.

[12] P. H. Swain, H. Hauska, **1977**. The decision tree classifier: Design and potential. *IEEE Transactions on Geoscience Electronics*, **1977,** *15*, 3, pp.142-147.

[13] H. Ramchoun, M. A. J. Idrissi, Y. Ghanou, M. Ettaouil, **2016**. Multilayer Perceptron: Architecture Optimization and Training. *IJIMAI*, **2016,** *4*, 1, pp.26-30.

[14] K. Esbensen, B. Swarbrick, F. Westad, P. J. Whitcomb, M.J. Anderson. *Multivariate data analysis: an introduction to multivariate analysis, process analytical technology and quality by design, 6th ed.* Camo Software AS, Magnolia TX, USA, **2018** pp. 166-170.

[15] G. W. Flake, S. Lawrence, **2002**. Efficient SVM regression training with SMO. *Machine Learning*, **2002,** *46*, 1-3, pp.271-290.

[16] V. Vapnik. *The nature of statistical learning theory*. Springer-Verlag, New York, USA, **2000,** pp. 131-167.

[17] M. T. Puth, M. Neuhäuser, G. D. Ruxton, **2014**. Effective use of Pearson's product–moment correlation coefficient. *Animal behaviour*, **2014,** *93*, pp.183-189.

[18] J. Kettle, V. Stoichkov, D. Kumar, M. Corazza, S. A. Gevorgyan. F. C. Krebs, **2017**. Using ISOS consensus test protocols for development of quantitative life test models in ageing of organic solar cells. *Solar Energy Materials and Solar Cells*, **2017,** *167*, pp.53-59.

[19] A. O. Oliynyk, J. M. Buriak, **2019**. Virtual Issue on Machine-Learning Discoveries in Materials Science. *Chemistry of Materials*, **2019,** *31*, 10, pp. 8243-8247.

[20] V. Tshitoyan, J. Dagdelen, L. Weston, A. Dunn, Z. Rong, O. Kononova, K. A. Persson, G. Ceder, A. Jain, **2019**. Unsupervised word embeddings capture latent knowledge from materials. science literature. *Nature*, **2019,** *571*, 7763, pp.95-98.

[21] Ç. Odabaşı, R. Yıldırım, **2019**. Performance analysis of perovskite solar cells in 2013–2018 using machine-learning tools. *Nano Energy*, **2019**, *56*, pp.770-791.

[22] J. Li, B. Pradhan, S. Gaur, J. Thomas, **2019**. Perovskite Solar Cells: Predictions and Strategies Learned from Machine Learning to Develop High-Performing Perovskite Solar Cells (Adv. Energy Mater. 46/2019). *Advanced Energy Materials*, **2019,** *9*, 46, p.1970181.

[23] Ç. Odabaşı, R. Yıldırım, **2020**. Machine learning analysis on stability of perovskite solar cells. *Solar Energy Materials and Solar Cells*, **2020,** *205*, p.110284.

[24] A. Tiihonen, K. Miettunen, J. Halme, S. Lepikko, A. Poskela, P.D. Lund, **2018**. Critical analysis on the quality of stability studies of perovskite and dye solar cells. *Energy & Environmental Science*, **2018,** *11*, 4, pp.730-738.

[25] M. Saliba, J. P. Correa-Baena, C. M. Wolff, M. Stolterfoht, N. Phung, S. Albrecht, D. Neher, A. Abate, **2018**. How to Make over 20% Efficient Perovskite Solar Cells in Regular (n–i–p) and Inverted (p–i–n) Architectures. *Chemistry of Materials*, **2018,** *30*, 13, pp.4193-4201.

[26] M. V. Khenkin, E. A. Katz, A. Abate, G. Bardizza, J. J. Berry, C. Brabec, F. Brunetti, V. Bulović, Q. Burlingame, A. Di Carlo, R. Cheacharoen, **2020**. Consensus statement for stability assessment and reporting for perovskite photovoltaics based on ISOS procedures. *Nature Energy*, **2020,** *5*, 1, pp.35-49.

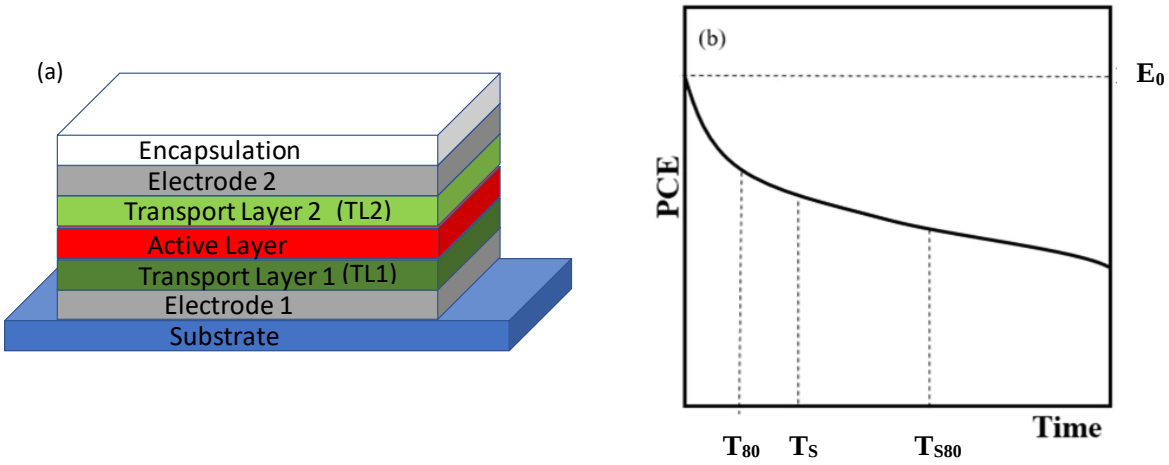

FIG.1: (a) Device architecture employed in dataset [4]. (b) Schematic of typical OPV degradation curve, illustrating key lifetime metrics employed in dataset.

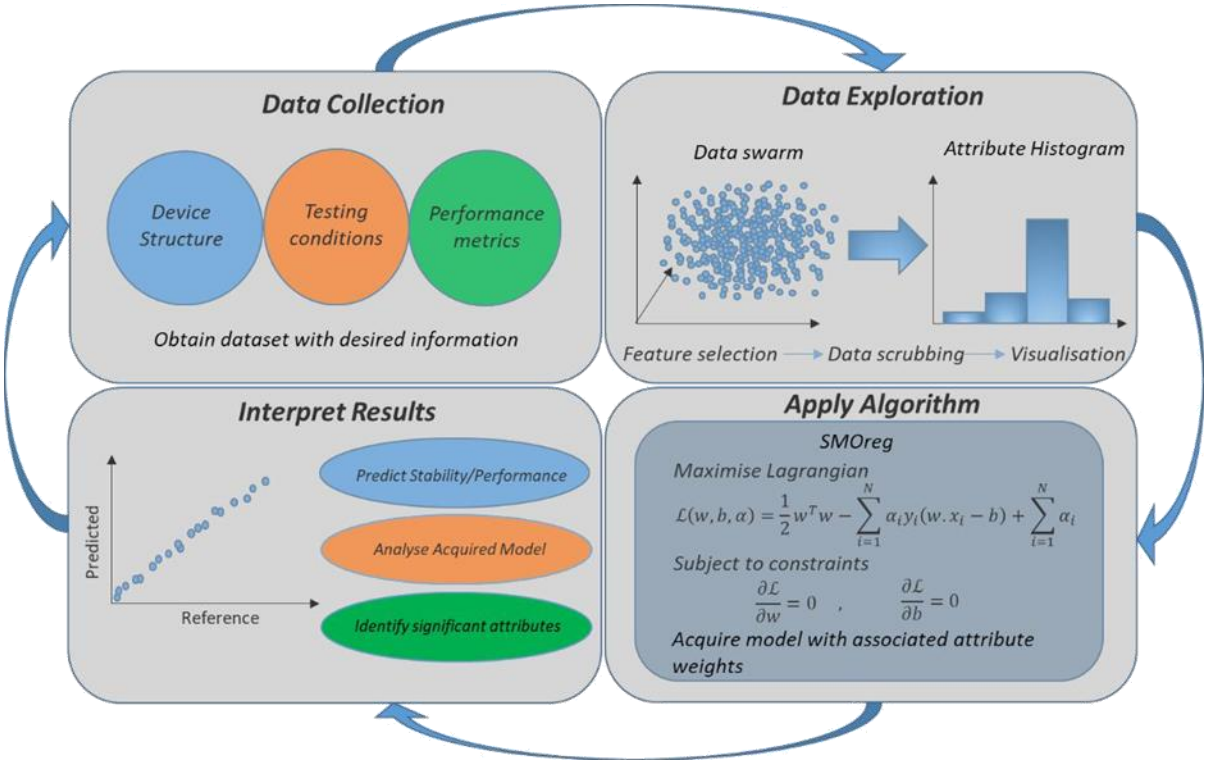

FIG.2: Schematic illustration of data science life-cycle.

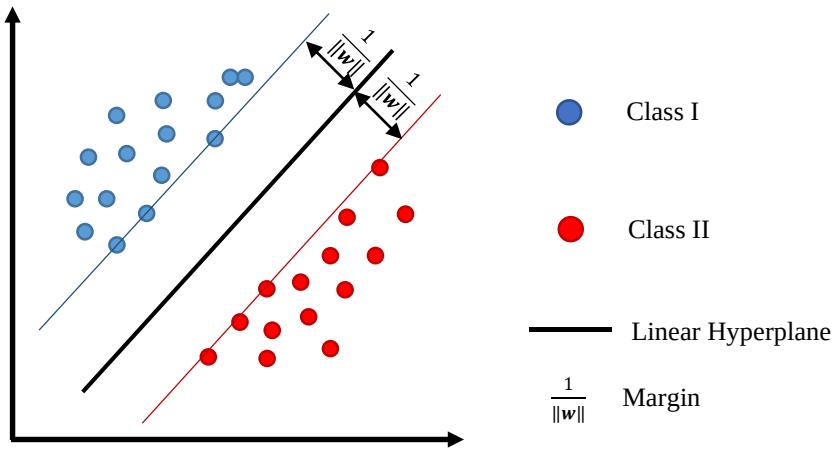

FIG.3: Schematic of classification and regression during SMOreg.

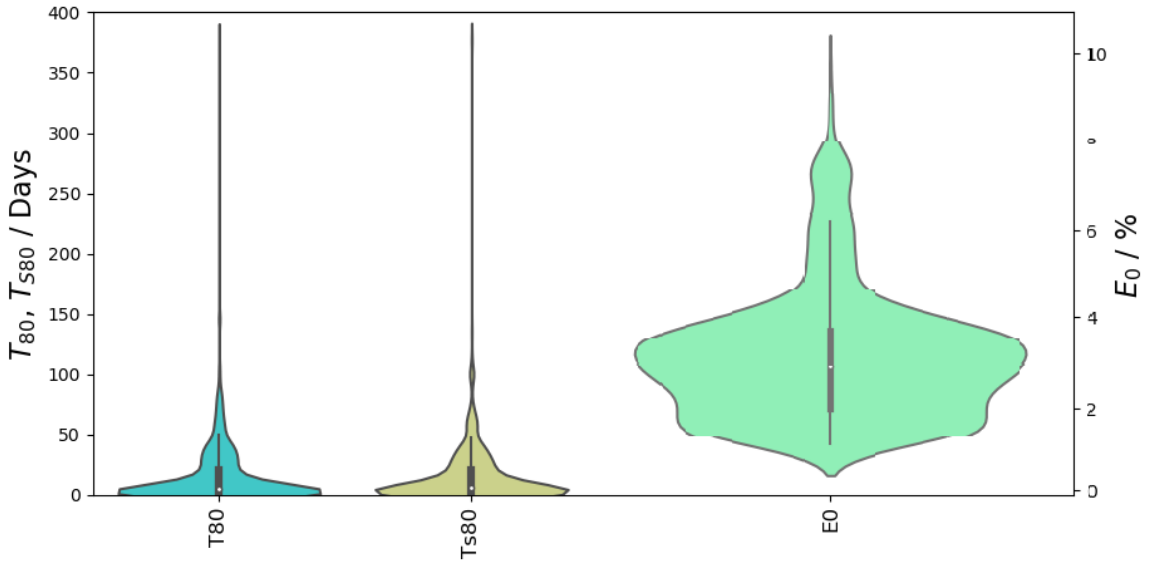

FIG.4: Categorical distribution of T_{80}, T_{S80} and E_0 for the full dataset.

	T_{80}	T_{S80}	E_0
Mean / Days, %	14.9	18.7	2.7
Median / Days, %	4.6	7.1	2.6
Max / Days, %	379.2	379.2	9.1
Range / Days, %	378.8	379.2	9.1

TABLE.1: Statistics for T_{80}, T_{S80} and E_0 distributions.

Metric	**All data**	**Only ISOS-L data**	**Only ISOS-D data**
Correlation coefficient	0.739	0.819	0.767
Mean Absolute Error (days)	9.31	7.64	12.1
Root Mean Squared Error (days)	16.2	20.9	23.5
Relative Absolute Error	59.4%	44.1%	52.9%
Root Relative Squared Error	69.3%	59.7%	65.1%
Number of Instances	1149	155	489

TABLE.2: T_{80} fit parameters for SMOreg algorithm applied to full dataset, ISOS-L and ISOS-D, based upon the training set.

Entire Dataset				**ISOS-L**				**ISOS-D**			
Best 10 Attributes		Worst 10 Attributes		**Best 10 Attributes**		Worst 10 Attributes		**Best 10 Attributes**		Worst 10 Attributes	
Name	Weight	Name	Weight	Name	Weight	Name	Weight	Name	Weight	Name	Weight
TL1 = NDP2 doped PV-TPD	0.2974	ISOS-T-3	-0.1033	Configuration = Tandem	0.3826	Conditions = Inert	-0.2431	Active2 = PTCDI	0.1954	TL1 = PEDOT:PSS	-0.1926
Active2 = PTCDI	0.0823	ISOS-L-1	-0.1006	Active1 = ZnPc	0.1873	Configuration = Normal	-0.1928	Configuration = Tandem	0.1375	Active2 = None	-0.1466
Active1 = PBDTTT-c	0.0571	ISOS-L-2	-0.0982	Light Type = Xenon	0.1138	Configuration = Inverted	-0.1897	TL2 = BPhen	0.0987	Configuration = Normal	-0.0864
Electrode1 = FTO	0.0554	ISOS not compatible	-0.0902	Light Type = Halogen	0.1041	Active1 = Unknown	-0.0576	Electrode2 = Au	0.0913	Temperature	-0.0745
TL1 = ZnO (Spray coated)	0.0483	ISOS-D-2	-0.0891	Light Type = FL	0.0935	Active1 = P3HT	-0.0416	Electrode1 = Ag Grid	0.0737	TL2 = PEDOT:PSS	-0.0647
Active1 = ZnPc	0.0308	ISOS-L-3	-0.0884	Active2 = PCBM (Slot die)	0.0167	Active1 = Other	-0.0374	Substrate = PEN	0.0647	TL2 = V_2O_5	-0.0594
Active2 = ICBA	0.0267	ISOS-D-3	-0.088	Electrode1 = Ag grid (flexible)	0.0163	TL1 = PEDOT:PSS	-0.0366	Encapsulation = Glass	0.0508	Active1 = PTB7	-0.0539
TL1 = NDP2(Noveled) doped DiNPB	0.0267	ISOS-D-1	-0.087	Substrate = Glass	0.0134	Active2 = C60	-0.0215	Encapsulation Adhesive = Epoxy	0.0416	Configuration = Inverted	-0.0511
Substrate = PEN	0.0214	ISOS-O-1	-0.0431	Type = Module	0.0095	Temperature	-0.0185	Encapsulation Adhesive = UV curable	0.0392	Active1 = PBTTTz-4	-0.0485
Light Type = LED	0.0205	Intensity	-0.0356	Electrode2 = Ag	0.0087	TL1 = None	-0.0181	Active1 = Pentacene	0.0295	Active1 = CuPc	-0.044

TABLE.3: Best 10 attributes and worst 10 attributes for T_{80} enhancement for entire dataset, ISOS-L and ISOS-D Yellow shading represents testing conditions. Green represents architectural components.

Data set	Electrode 1	TL 1	Active 1	Active 2	TL 2	Electrode 2
Full	FTO	NDP2(Novaled) doped PV-TPD	ZnPc	ICBA	ZnO-np	Al
	Cr/Al	ZnO- spray-coated	PBDTTT-c	PCBM	ZnO	Ag
	Ag	NDP2(Novaled) doped DiNPB	PC-TBT-TQ	C60	PEDOT:PSS	Au
ISOS-L	Ag grid flexible	BF-DPB	ZnPc	PCBM slot-dye	BPhen	Al
	Other	PEDOT:PSS	Other	PCBM	BaF2	Ag
	ITO	Unknown	P3HT	PCBM - 71	PEDOT:PSS	Ag grid
ISOS-D	Ag grid	PEDOT:PSS	PECz-DTQx	PTCDI	BPhen	Au
	ITO	ZnO – spray-coated	MDMO-PPV	PCBM slot-dye	AlQ3	Ag
	Other	TiOx	ZnPc	ICBA	BCP	Other

TABLE.4 Top three attributes for each layer in device architecture for the full dataset, ISOS-L and ISOS-D.

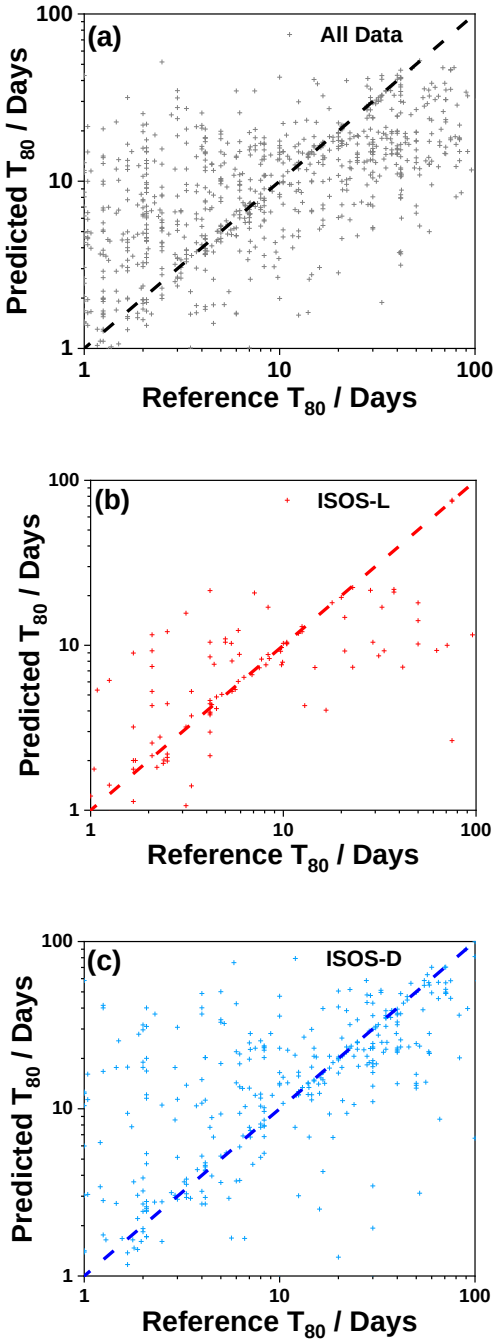

FIG.5: Predicted T_{80} lifetimes plotted as a function of reference T_{80} lifetimes for (a) the full dataset, (b) ISOS-L testing and (c) ISOS-D testing. The dashed line represents a perfect correlation of 1:1. Please not, the data is plotted on a logarithmic scale.

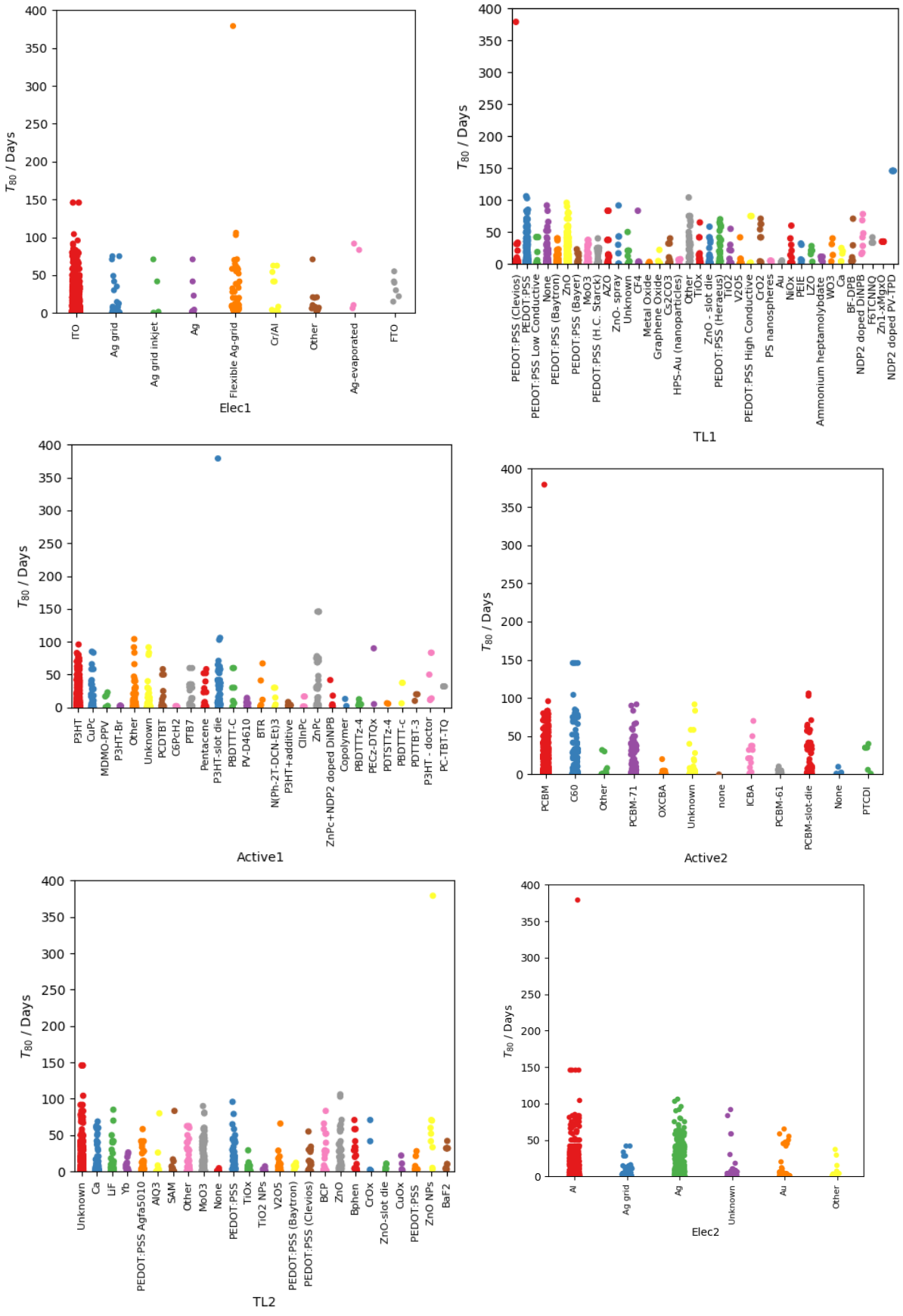

FIG.6: Distribution of T_{80} lifetimes for each class present in dataset.

Metric	Full Dataset
Correlation Coefficient	0.739
Mean Absolute Error / %	0.605
Root Mean Squared Error / %	0.939
Relative Absolute Error	55.6%
Root Relative Squared Error	63.9%
Number of Instances	1347

TABLE.5: E0 fit parameters for SMOreg algorithm applied to full dataset based on training.

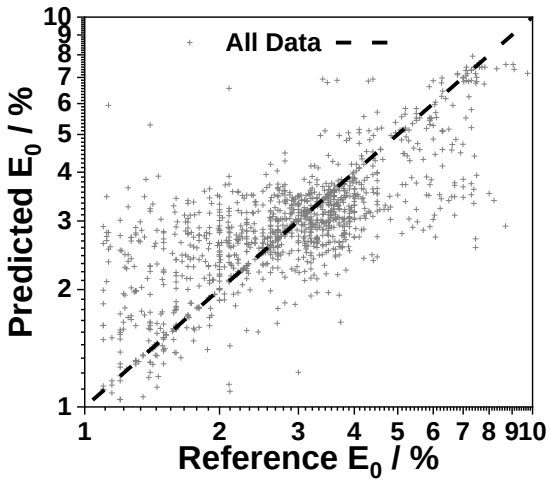

FIG.7 Predicted E_0 performance plotted as a function of reference E_0 performance for the full dataset, plotted on a logarithmic scale. The dashed line represents a perfect correlation of 1:1.

Entire Dataset			
Best 10 Attributes		Worst 10 Attributes	
Name	Weight	Name	Weight
Active1 = PTB7	0.3572	TL1 = MeO-TPD	-0.1689
Configuration = Tandem	0.2315	Active1 = PBbTTT-T	-0.1551
Active1 = PBDTTT-c	0.2097	TL1 = NDP2(Noveled) doped PV-TPD	-0.1522
Active1 = PCDTBT	0.1769	Active1 = CuPc	-0.1372
Active2 = ICBA	0.1581	TL1 = AZO	-0.1262
Active2 = OXCBA	0.158	Configuration = Inverted	-0.1221
Electrode1 = Ag	0.1445	Intensity	-0.1185
TL1 = Graphene oxide	0.1042	Active1 = P3HT (Slot die)	-0.1133
TL2 = V_2O_5	0.0889	Electrode1 = Ag grid (Inkjet)	-0.1012
Electrode2 = Pt	0.0788	Active2 = None	-0.0989

TABLE.6: Best 10 attributes and worst 10 attributes for E_0 enhancement for entire dataset.